Does HEP have a "Titanic" Problem

Frank Würthwein
UCSD/SDSC

October 6th 2020

I was asked by the conveners to be provocative

You will get to judge whether I accomplished my mission.

Commonality to the Titanic

- HEP is heavily dominated by large experiments
- Large experiments tend to produce large monolithic platforms that are maintained by large teams to improve “ease of use”
  - Large experiments tend to be incapable of making quick turns

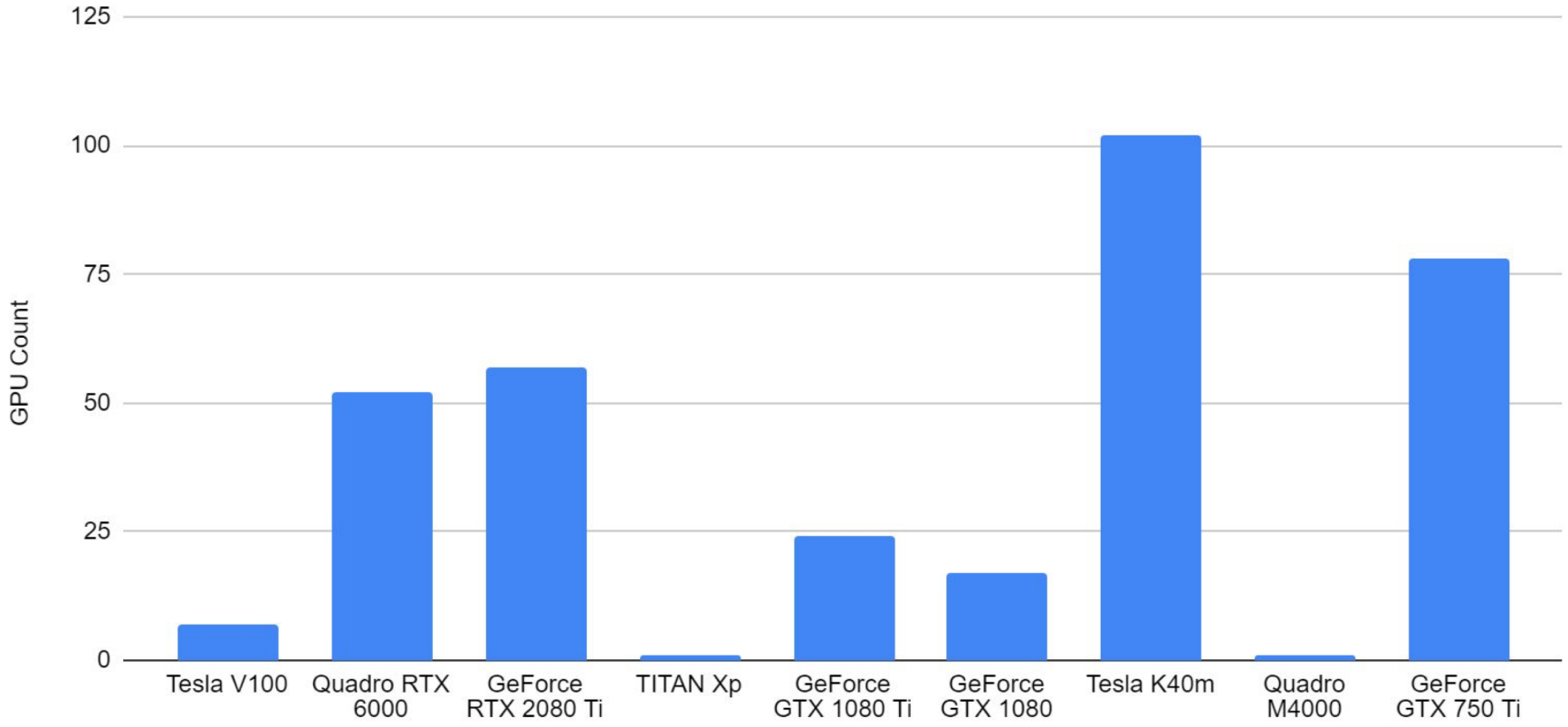
Should we consider becoming more agile by providing less?

Some Examples

I'm picking examples from OSG because that's what I know best.
The common thread is that your random biologist
does more with less effort, and adjusts faster to new technology
than the large experiments in HEP.

Why?

GPU Computing on OSG



Wide variety of GPUs available to anybody who wants it.
Fastest growing segment of hardware available via OSG

Lot's of individual PIs from many fields use them, but HEP largely absent.
(exception is IceCube, of course)

How is it supported?

1. GPUs are now widely available in the OSG open pool
2. Match jobs against attributes/capabilities, not specific models
3. Use provided Singularity images to get started

Documentation:

https://support.opensciencegrid.org/support/solutions/articles/5000653025-gpu-jobs

Available Containers:

https://support.opensciencegrid.org/support/solutions/articles/12000073449-available-containers-list

Questions?

support@osgconnect.net

Link to talk at OAG AHM for more details

Machine Learning

- Introduced to OSG by individual PI from bioinformatics.
- Dominant user community is from life sciences, plus some quantum chemistry.

Predicting Protein Activity via ML

ML accelerated Molecular Modelling

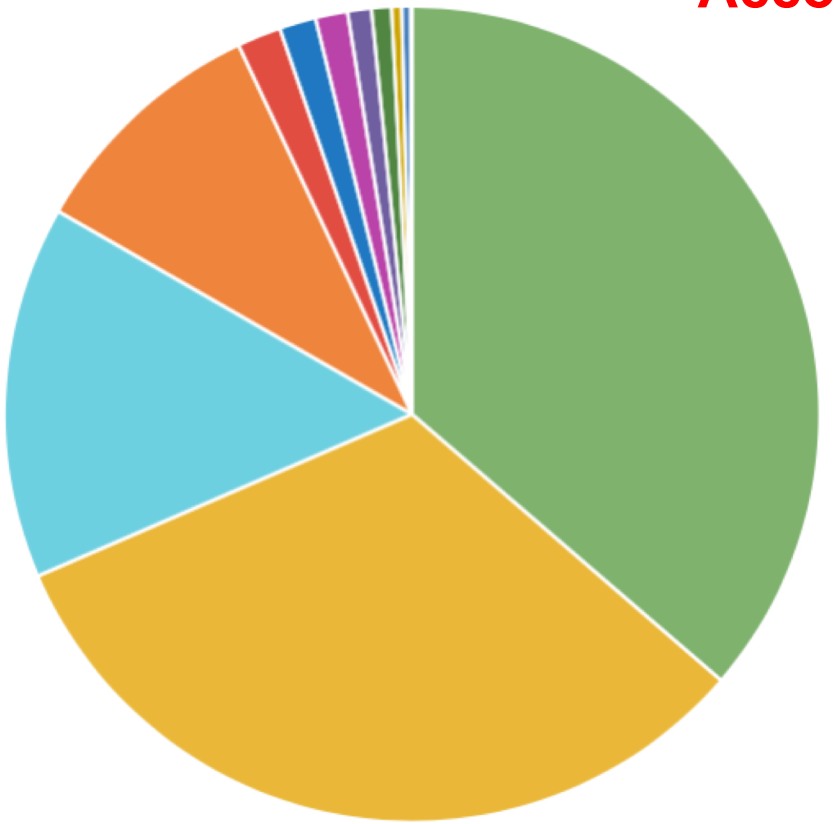
How to for ML on OSG

Why is there so little ML from HEP on OSG ?

OSG Connect Community

Core Hours by Field of Science

Accounting for the last year



	total
Biological Sciences	93.4 Mil
Physics	82.8 Mil
Astronomy	38.0 Mil
Chemistry	25.0 Mil
Engineering	4.47 Mil
Agricultural Sciences	3.73 Mil
Health	3.29 Mil
Integrative Activities	2.425 Mil
Other	2.022 Mil
Mathematics	1.024 Mil
Computer Sciences	891 K
Economics	130.9 K
Earth and Ocean Sciences	25.4 K
Education	0.3

More details in Lauren's talk in this session

What's common?

- All of these sciences share a minimal API
  - Submission to a standard batch system
  - Standard runtime environment & containers
- This way any innovation for anybody becomes immediately available to all.

How is HEP different?

- We insulate our community via CRAB/Panda/Dirac/FIFE/…. from a common API shared by all of science.
- We build elaborate software environments that nobody else on the planet uses.
- All in the name of “ease of use”
- We then train our communities on our idiosyncratic platforms rather than making them computer literate on general APIs and environments common to all of science.